

VU Research Portal

De kunst van het voorspellen in beweging

den Butter, F.A.G.

published in
Econometrie in Beweging
1997

[Link to publication in VU Research Portal](#)

citation for published version (APA)
den Butter, F. A. G. (1997). De kunst van het voorspellen in beweging. In G. van der Laan (Ed.), *Econometrie in Beweging* (pp. 1-12). Vrije Universiteit.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

De kunst van het voorspellen in beweging

Frank A.G. den Butter*

1. Inleiding

De zucht naar kennis over de toekomst vormt voor vele wetenschappers een bron van inspiratie. Dit geldt zeker voor economen. Immers, in de economie is het om verschillende redenen van belang om goed de toekomst te kunnen voorspellen. Zo zijn voorspellingen over de economische ontwikkeling nodig als achtergrondinformatie voor ondernemers en beleidmakers bij het nemen van strategische beslissingen. Daarbij valt overigens te betwijfelen of iemand die echt goed de economische toekomst kan voorspellen, de voorspellingen openbaar moet maken, omdat de beslissingen op basis van deze voorspellingen immers voor de goede voorspeller zeer winstgevend kunnen zijn.

Hierin verschillen economische voorspellingen van weersvoorspellingen. Trouwens, weersvoorspellingen hebben meestal op een kortere termijn betrekking dan economische voorspellingen. Bovendien wordt er, zoals de Engelse econometrist David Hendry heeft opgemerkt (zie Granger, 1996), althans door de Britse overheid, nog een asymmetrisch beleid gevoerd ten aanzien van economische voorspellers en weersvoorspellers. Wanneer de weersvoorspellers een verwoestende storm niet goed hebben voorspeld, krijgen ze grotere computers om het de volgende keer beter te doen, terwijl de economische voorspellers na een verkeerde voorspelling juist op hun budget worden gekort.

Behalve als hulpmiddel bij strategische beslissingen zijn in de economische wetenschap voorspellingen nog om een tweede reden van belang. De kwaliteit van de voorspellingen vormt namelijk een maatstaf voor de realiteitswaarde van de economische modellen waarmee de voorspellingen zijn verkregen, en daarmee ook voor de realiteitswaarde van de economische theorieën waarvan deze modellen een formele weergave zijn. Deze rol van voorspellingen in de economische wetenschap staat centraal in het proefschrift van Merkies (1972), waarin hij de voorspelkwaliteit afmeet aan het voorspellingsinterval en deze als selectie criterium hanteert bij de modelkeuze.

* Hoogleraar Algemene Economie, Onderzoeksgroep Toegepaste Arbeidseconomie (ALERT), Vrije Universiteit, Amsterdam. In 1971 en 1972 heeft de auteur als kandidaatsassistent rekenwerk voor het proefschrift van Merkies verricht en in de periode 1995/1997 is hij tijdens het decanaat van Merkies vice-decaan geweest.

Tegen deze achtergrond bevat dit artikel enige reflecties over het gebruik van modellen bij economische voorspellingen. Dit gebruik vertoont gelijkenis met het gebruik van de kristallen bol door een waarzegster. Net zoals de waarzegster de kristallen bol als instrument (medium) voor het doen van voorspellingen hanteert, zo is ook het model voor de economische voorspeller een onontbeerlijk hulpmiddel. Maar alleen de waarzegster kan de spiegelingen in de bol goed interpreteren en deze omvormen tot voorspellingen. Evenzo kleurt ook de professionele voorspeller de uitkomsten van het model met zijn eigen inzichten. Als zodanig is het voorspellen op te vatten als een kunde, waarbij het model het wetenschappelijke aspect van de voorspelprocedure representeert. Aldus bestaat een onderscheid tussen *de wetenschap van het voorspellen* en *de kunst van het voorspellen*.

De gedachte van Merkies om modellen op basis van hun voorspelkwaiteit te selecteren betreft de wetenschap van het voorspellen. In dit artikel geef ik eerst een impressie van de wijze waarop deze wetenschap van het voorspellen in beweging is. Daartoe worden in paragraaf 2 de ideeën van Merkies over modelselectie besproken en vergeleken met andere criteria om de werkelijkheidswaarde van economische modelspecificaties te beoordelen. Vervolgens gaat paragraaf 3 over tijdreeksvoorspellingen, een onderwerp waarvoor Merkies indertijd mijn belangstelling heeft gewekt. In paragraaf 4 komt de kunst van het voorspellen aan de orde. Dit is een bewerking van een eerder artikel hierover (Den Butter, 1992; zie ook Compaijen en Den Butter, 1996). Deze bijdrage aan de afscheidsbundel wordt afgesloten met de slotbeschouwing in paragraaf 5.

2. Modelkeuze en voorspellingen

Een belangrijke karakteristiek van de economische wetenschap is dat economische verschijnselen zich niet ondubbelzinnig door een bepaalde theorie en het daarop gebaseerde formele model laten beschrijven. Anders gezegd, er bestaat in de economie geen consensus over de onderliggende theorie. Voor de verklaring van de meeste verschijnselen zijn meerdere concurrerende theorieën en modellen beschikbaar. Zeker wanneer deze verschillende theorieën voor hetzelfde verschijnsel tot afwijkende beleidsaanbevelingen leiden, richt het economische discours zich op de werkelijkheidswaarde van deze alternatieve theorieën. In die zin verschilt de economische wetenschap ten principale van bij voorbeeld de natuurkunde, waaraan de economische wetenschap zich - naar mijn mening ten onrechte - nog wel eens graag spiegelt. Terwijl er in de natuurkunde een sequentieel komen en gaan van theorieën is - de nieuwe theorie vervangt de oude theorie waarmee de oude theorie overbodig is - is het naast elkaar bestaan van verschillende theorieën over het zelfde verschijnsel een essentieel kenmerk van de economische wetenschap. Het betekent dat het vraagstuk van de modelkeuze binnen de economische wetenschappen een eigen behandeling behoeft. Hierbij kunnen de selectiecriteria niet aan andere wetenschappen worden ontleend. Deze specifieke functie van de modelselectie binnen de economie tekent de methodiek van modelselectie die Merkies in zijn proefschrift heeft ontwikkeld.

Binnen de modelselectie in de economie speelt momenteel het beginsel van de overkoepeling ('encompassing') een belangrijke rol. Hierbij wordt voor twee of meerdere naast elkaar staande

theorieën voor een bepaald verschijnsel een model opgesteld, dat de elementen van de verschillende theorieën in zich verenigt. Deze theorieën zijn dan ieder een specifiek geval van het overkoepelende model. Een voorbeeld is het sequentiële loon- en werkgelegenheids-onderhandelingsmodel van Manning (1987). Dit model vormt een overkoepeling van de monopolie vakbondstheorie, het 'right-to-manage model' en de efficiënte onderhandelingstheorie: drie theorieën die ieder op hun eigen wijze het proces van loonvorming in de onderhandeling tussen vakbond en werkgevers beschrijven. Bij empirische toetsing kan een dergelijk overkoepelend model inzicht - maar vaak geen uitsluitsel - bieden over de vraag welke van de achterliggende theorieën de meeste werkelijkheidswaarde heeft.

In de selectieprocedure van Merkies komt dát model als juiste specificatie uit de bus, dat voor de te verklaren grootheid - of te verklaren grootheden (maar twee is in dit geval al heel veel) - het kleinste voorspellingsinterval oplevert. Het gaat daarbij om een volledig gemechaniseerde selectieprocedure, waarbij een aantal elementen in de selectie op heuristische wijze is ingebouwd. De procedure start met het kleinste gemene veelvoud van mogelijke verklarende variabelen die de verschillende theorieën aandragen. Dit is bepalend voor de verzameling van toegelaten modellen, waarbij naast theoretische overwegingen ook technische overwegingen en de beschikbaarheid van gegevens een rol spelen. De algoritmische selectieprocedure kiest uiteindelijk uit de verzameling van toegelaten modellen het model met het kleinste voorspellingsinterval. In deze zin toont de selectieprocedure van Merkies verwantschap met het beginsel van de overkoepeling en is als een voorloper 'avant la lettre' te beschouwen van het beginsel van 'van algemeen naar specifiek' (from general to specific) in de moderne specificatie-analyse van tijdreeksmodellen. Opmerkelijk is ook dat de methodiek van Merkies de zelfde, nogal ondergeschikte rol aan de theorie toekent als later Sims (1980) in zijn methodiek van de modellering van macro-economische tijdreeksen aan de hand van vector-autoregressieve (VAR)-stelsels. Vermeldenswaard is tevens dat deze selectieprocedure van Merkies berust op de omvang van de voorspellingsintervallen en dat voor de modelkeuze dus geen gegevens over de uitkomsten van de voorspellingen nodig zijn. Tegenwoordig zijn bij de modelselectie en specificatieanalyse van tijdreeksmodellen vooral toetsen in gebruik die bezien in hoeverre de realisatie van voorspellingen binnen het voorspellingsinterval ligt.

Bij wijze van anekdote zij vermeld dat de selectieprocedure van Merkies om uit de toegelaten verzameling modellen het model met het kleinste voorspellingsinterval te selecteren, nogal rekenintensief was, althans in het begin van de jaren '70 toen de berekeningen zijn gemaakt. In zijn proefschrift heeft Merkies daarom slechts voorbeelden van de uitkomst van de selectieprocedure getoond voor een verzameling modellen met een zeer beperkt aantal verklarende variabelen. Ofschoon Merkies wel op theoretische gronden het naast elkaar opnemen van bepaalde verklarende variabelen uitsluit, en aldus bereid is iets meer identificerende restricties aan de theorie te ontleen dan Sims, neemt het aantal te beschouwen modellen toch sterk toe met het aantal verklarende variabelen. Dit leverde Merkies tijdens de promotieplechtigheid een kritische vraag van De Wolff op of de procedure van modelkeuze wel reken-technisch uitvoerbaar was, indien deze werd toegepast op modellen met een iets groter aantal mogelijke verklarende variabelen. Om de grenzen van de rekenmogelijkheden, die de ELX8 van het Mathematisch Centrum indertijd bood, aan te geven, heeft Merkies (1973) in de Engelse versie van zijn proefschrift een appendix (blz. 130-132) opgenomen waarin een rekenvoorbeeld met acht verklarende variabelen wordt gegeven, met in

totaal 255 verschillende toegelaten modellen. De computer bleek hierbij toch nog binnen de 5 minuten tot een oplossing te komen.

Het proefschrift van Merkies bevat zeker niet het laatste woord op het gebied van de modelselectie. Dit geldt in retrospect maar waarschijnlijk was dit ook wel op voorhand aan de schrijver duidelijk. In een uitgebreide en van waardering getuigende boekbespreking van het proefschrift noemt Kloek (1974) een aantal bedenkingen tegen de procedure van Merkies, maar geeft hij ook een aantal interessante opties voor een vervolgstudie. Zo geldt bij de berekening van het voorspellingsinterval impliciet de veronderstelling dat het model qua stochastiek juist gespecificeerd is. Merkies verwerpt weliswaar op voorhand die modellen uit de verzameling toegelaten modellen, waar statistische toetsen op specificatiefouten duiden, maar hij houdt niet, zoals in de analyse van Theil, rekening met de mogelijkheid dat andere modellen correct zijn zodat de selectiemaatstaf moet worden afgeleid onder de hypothese dat een ander model inderdaad het correcte is. Volgens Kloek kan men bewijzen dat het criterium van Merkies onder betrekkelijk zwakke voorwaarden bij voldoende grote steekproeven met grote kans het goede model aanwijst. Deze eigenschappen heeft het criterium echter gemeen met de correlatiecoëfficiënt, die in de praktijk veelvuldig gebruikt wordt als selectiecriterium. Interessant is daarom, aldus Kloek, juist de vraag of het criterium van Merkies bij kleine steekproeven beter voldoet dan de correlatiecoëfficiënt. Pas dan is het immers relevant om dit toch flink ingewikkelde en rekenintensieve selectiecriterium te gebruiken.

Meer in het algemeen kan men zich met Kloek afvragen of het zinvol is te veronderstellen dat één model uit de verzameling het juiste is en dat dit model dan gebruikt moet worden voor een puntvoorspelling. De theorie van de combinatie van voorspellingen leert immers dat een gecombineerde voorspelling op basis van meerdere modellen waarbij ieder model additionele, relevante informatie over de te voorspellen grootte bevat, beter voldoet dan ieder van de voorspellingen met de afzonderlijke modellen. Dit houdt verband met de methodologische kwestie of het in de tijdreeksanalyse zowieso zinvol is uit te gaan van het bestaan van een 'gegevensgenererend proces' (data generating process). Dit paradigma vormt het uitgangspunt van de alom nagevolgde en toegepaste modelspecificatie en -selectiemethodiek van Hendry en de zijnen. Het is echter weinig waarschijnlijk dat in werkelijkheid tijdreeksgegevens volgens een dergelijk proces tot stand komen. In dat geval wordt de speurtocht naar het juiste model een zoektocht naar de Steen der Wijzen en dus pure alchemie (zie ook Keuzenkamp, 1995). Toegegeven zij echter dat Merkies in zijn selectieprocedure niet op zoek was naar het juiste model maar naar het best voorspellende model.

Tot slot van zijn boekbespreking meldt Kloek dat de vraag blijft bestaan of het verstandig is te vertrouwen op mechanische selectiemethodes in het geval dat relatief weinig waarnemingen ter beschikking staan. Mijn twijfel is op dit punt groter dan die van Kloek. Ook al zijn er wel veel waarnemingen beschikbaar, toch acht ik de economische inzichten en de oordeelmatige informatie buiten het model om zodanig belangrijk dat de mechanische modelselectie hooguit als een hulpmiddel bij het doen van voorspellingen over de economie kan worden beschouwd. In paragraaf 4 wordt dit verder toegelicht.

Dat het laatste woord over het gebruik van voorspelcriteria voor modelselectie nog niet gesproken is blijkt uit de artikelen van Van Gameren, Molenaar en Paap (1993) en van Christoffersen en Diebold (1996). Terwijl Christoffersen en Diebold het gebruik van niet-symmetrische verliesfuncties bij de modelselectie voorstaan, beperken Van Gameren *c.s.* zich bij hun evaluatie van modelselectiecriteria tot een aantal aan de literatuur ontleende criteria waarin de voorspelfouten, d.w.z. de verschillen tussen de met het model voorspelde waarden en de realisaties, in beginsel symmetrisch wordt behandeld. Opgemerkt zij dat, anders dan bij Merkies, voor alle hier gehanteerde voorspelcriteria de realisaties van de voorspelde waarden nodig zijn. Van Gameren *c.s.* nemen dan ook niet het modelselectie criterium van Merkies, dat overigens aan Wiener is ontleend, op in hun lijst met onderzochte criteria. Wellicht is dit ook het gevolg van het feit dat, zoals zo vaak, een nieuwe generatie relevant werk van een voorgaande generatie over het hoofd ziet. Van Gameren *c.s.* beperken hun onderzoek naar de kwaliteit van modelselectiecriteria tot univariate autoregressieve tijdreeksmodellen (zie ook de volgende paragraaf). Een opmerkelijke uitkomst van deze simulatiestudie is dat de modelselectie beter kan worden gebaseerd op gegevens binnen de steekproefperiode dan op voorspelcriteria toegepast op extra waarnemingen, waarbij dus, zoals gezegd, realisaties buiten de steekproef nodig zijn. Misschien betekent dit dat ook het modelselectie criterium van Merkies goed uit deze vergelijkende analyse naar voren zou zijn gekomen.

3. Tijdreeksvoorspellingen

Merkies heeft voor zijn studie naar de modelselectiecriteria vooral inspiratie geput uit zijn voorgaande werk in de modelbouw bij het CPB. De modelselectie via voorspelcriteria is echter met name relevant in het geval van tijdreeksmodellen. In dat geval gaat het om onvoorwaardelijke voorspellingen, waarbij ontwikkelingen in het heden en het verleden bepalend zijn voor de ontwikkeling in de toekomst. Voorspellingen met behulp van macro-economische structuurmodellen zijn daarentegen vaak conditioneel, aangezien men voor de voorspellingen waarden nodig heeft voor de exogene grootheden in het model. Men moet dan eerst over deze exogenen toekomstverwachtingen vormen voordat men de eigenlijke voorspellingen met behulp van het model kan berekenen. Aanpassingsvertragingen, gespecificeerd via vertraagde endogenen, zijn in de structuurmodellen essentieel om tot een voorspelling te kunnen komen. In feite introduceren deze vertraagde endogenen in het macro-economische structuurmodel elementen van een tijdreeksmodel. Deze vertraagde endogenen krijgen dan ook in de modelselectieprocedure van Merkies een aparte behandeling.

Mij bracht dit indertijd op de gedachte om het onderzoek naar de kwaliteit en de mogelijke kwaliteitsverbetering van modelvoorspellingen toe te spitsen op autoregressieve tijdreeksmodellen. Het meest eenvoudige voorbeeld van zo'n model is een autoregressief proces van de eerste orde, ofwel een AR(1)-model in de terminologie van Box en Jenkins (1970). Toevalligerwijs was het baanbrekende werk van Box en Jenkins net verschenen in de periode dat Merkies zijn proefschrift aan het afronden was. Het idee dat ik indertijd op basis van gesprekken met Merkies in mijn scriptie heb uitgewerkt en dat Merkies ook nog in zijn proefschrift heeft vermeld, was dat met het autoregressieve model wellicht een andere voorspelmethodiek nodig is om meer perioden vooruit

te voorspellen dan wanneer het om één periode vooruit voorspellen gaat (zie Den Butter, 1975). Immers, de parameterschattingen van de tijdreeksmodellen, en eveneens van door regressie bepaalde economische structuurvergelijkingen, zijn - in beginsel - gebaseerd op minimering van de kwadratische afwijkingen van de residuen. In autoregressieve tijdreeksmodellen hebben deze residuen betrekking op de voorspelfout één periode vooruit binnen de waarnemingsperiode. De vraag dringt zich dan op of het niet beter is om, wanneer men twee perioden vooruit wil voorspellen, de kwadraten van de voorspelfout twee perioden vooruit te minimeren. Helaas bleek uit allerlei simulatie-experimenten die ik hierover indertijd onder gebruik van veel computertijd heb uitgevoerd, dat een aldus aangepaste schattingsmethodiek niet tot betere voorspellingen over meer perioden vooruit leidde. Dit idee bleek daarmee niet nuttig, maar het was wel geruststellend dat tijdreeksmodellen niet voor iedere extra vooruit te voorspellen periode opnieuw geschat behoeven te worden. Vermeldenswaard in dit verband is dat Van Gameren c.s. eveneens naar de kwaliteit van voorspelcriteria als selectiemethode hebben gekeken wanneer meerdere perioden worden voorspeld. Hun conclusie is dat modelselectie aan de hand van één-staps-voorspellingen betere resultaten geeft dan wanneer men voorspellingen voor meerdere perioden tegelijk beziet. Toch is dit idee van de samenhang tussen modelselectie en het voorspellen van meerdere perioden vooruit niet helemaal in de vergetelheid geraakt. Zo heeft Kool (1989) nog een door Merkies begeleid proefschrift verdedigd over modelselectie in multivariate tijdreeksmodellen op basis van criteria over het vooruit voorspellen van meerdere perioden. Hoewel Kool concludeert dat doorgeschoven VAR(h)-modellen, zoals hij deze noemt, een goed alternatief zijn voor gewone VAR-modellen bij het vooruit voorspellen van meerdere perioden, lijkt de winst in voorspelkwaliteit die met behulp van dergelijke technische verfijningen kan worden bereikt, relatief gering te zijn. Het is de vraag of zulke methoden lonend zijn en of voortzetting van het onderzoek in deze richting niet als 'penny wise but pound foolish' kan worden gekwalificeerd.

Meer in het algemeen speelt de tijdshorizon waarover voorspeld dient te worden natuurlijk wel een belangrijke rol bij de beoordeling van de kwaliteit van voorspellingen en bij de keuze van het model waarmee de voorspellingen worden uitgevoerd. Economische modellen lijken niet geschikt om erg ver, zeg 20 à 40 jaar vooruit te voorspellen. Of, om het in navolging van Granger (1996) meer nauwkeuriger te formuleren: voor voorspellingen ver vooruit in de economie kan men net zo goed heel eenvoudige modellen gebruiken, zoals het 'random walk' model met een deterministisch groeipad, dan meer ingewikkelde tijdreeksmodellen of economische structuurmodellen. Maar ook voor voorspellingen over een kortere periode kan de tijdshorizon van de voorspelling bepalend zijn voor de keuze van het model. Zo lijkt het niet zinvol om bij voorbeeld de kapitaalmarktrente over vier jaar te voorspellen met een tijdreeksmodel dat gebaseerd is op dagcijfers, ook al zijn deze gegevens beschikbaar en bevatten ze in theorie de meest uitvoerige informatie over de rentebewegingen. Om dit aspect te onderzoeken heb ik indertijd (Den Butter, 1976) voor een bepaalde reeks van rentegegevens de voorspelkwaliteit van een model gebaseerd op maandgegevens vergeleken met een kwartaalmodel. Hierbij bleek bij een voorspelling van twee of meer kwartalen vooruit het kwartaalmodel op basis van de variantie van de voorspelfout als criterium beter te voldoen dan het maandmodel. Opmerkelijk was verder dat de meest geschikte specificaties van de maand- en kwartaalmodellen niet met elkaar overeen kwamen. Het bleek dat de specificatie van het kwartaalmodel, die werd verkregen door intertemporele aggregatie van de meest geschikte

specificatie voor het maandmodel, minder goed voldeed dan een andere, meer eenvoudige specificatie voor het kwartaalmodel. Deze problematiek van verschillende specificaties bij intertemporele aggregatie toont dat het inderdaad weinig zinvol is om econometrische modelselectieprocedures en de methodiek van specificatieanalyse te baseren op het paradigma dat de waargenomen tijdreeks door een bepaald gegevensgenererend proces is voortgebracht. Immers, de vraag is dan over welk tijdsinterval dit proces de gegevens oplevert: per maand, per dag of zelfs per uur. Bij kleine tijdsintervallen leidt intertemporele aggregatie al spoedig tot gecompliceerde specificaties voor de tijdreeksmodellen, hetgeen weer in strijd is met de wens tot spaarzaamheid die alom bij de specificatie van tijdreeksmodellen als uitgangspunt geldt.

Al met al geloof ik niet dat technische verbeteringen in de tijdreeksanalyse ons nog veel winst in de kwaliteit van economische modelvoorspellingen kunnen opleveren. Dit geldt zeker op het terrein van de macro-economische voorspellingen. De wetenschap van het voorspellen heeft op dat punt zijn grenzen bereikt. Tekenend hiervoor is dat Granger (1996) zich in zijn recente overzichtsartikel in de eerste plaats richt op de wijze waarop de presentatie van de kwaliteit van de economische voorspellingen kan worden verbeterd en slechts in tweede instantie mogelijke technische verfraaiingen noemt. Vandaar dat het naar mijn mening aanbeveling verdient dat economische voorspellers hun energie veeleer stoppen in de kunst van het voorspellen dan in een uitbreiding van de technische kennis. In de volgende paragraaf wordt een aantal aspecten van deze kunst van het voorspellen besproken.

4. De kunst van het voorspellen

De voorgaande paragraaf gaat over het gebruik van modellen uit de tijdreeksanalyse voor economische voorspellingen. Het zijn echter niet deze 'mechanische' modellen maar veeleer de causale modellen die een wezenlijk *hulpmiddel* vormen bij *de kunst van het voorspellen*. Zoals gezegd zal, voor zover de economische voorspellingen voor verbetering vatbaar zijn, de winst vooral langs deze weg te bereiken zijn. Daarbij is wel een goede wisselwerking nodig tussen de wetenschappelijke voorspelmethode en de kunst van het voorspellen. Vandaar dat door instellingen, zoals het CPB, het IMF en de OESO, die op professionele basis macro-economische voorspellingen afleveren, vooral van causale modellen gebruik wordt gemaakt. De voornaamste wetenschappelijke uitdaging ligt hierbij in de uitbreiding van onze kennis over de gedragsmatige samenhangen in de economie. Economische theorie en empirie gaan daarbij hand in hand, waarbij tegenwoordig ook veel belang wordt gehecht aan empirische kennis die op basis van micro-economische gegevensbestanden is verkregen.

Voorheen is de gelijkenis gebruikt van de macro-economische voorspeller als waarzegster. Deze hanteert het model, of meer in het algemeen de economische en statistische methodologie, als een kristallen bol. Maar de kristallen bol zelf levert geen goede voorspelling op: de waarzegster is nodig om de spiegelingen in de bol te interpreteren. Hetzelfde geldt voor de macro-economische voorspellingen. De serieuze, in het beleid gebruikte voorspellingen komen nooit direct uit de computer rollen. De professionele voorspeller zal de voorspelling altijd van zijn eigen commentaar willen voorzien en willen vertellen hoe de voorspelling tot stand is gekomen. Vandaar dat de publikaties

van de voornoemde instellingen niet uitsluitend tabellen met voorspellingen bevatten, maar in hoofdzaak gevuld zijn met beschouwingen over de toekomstige economische ontwikkeling en met het aangeven van de veronderstellingen, of de voorwaarden, waarbij deze ontwikkelingen zullen optreden. Juist om deze redenen zijn de mechanische voorspelmethoden voor de macro-economische voorspelling minder zinvol dan de causale methoden: er valt geen verhaal bij te vertellen.

Daarentegen verschaft het gebruik van de causale macro-economische beleidsmodellen inzicht in de economische mechanismen waarop de voorspelling is gebaseerd. Bovendien biedt zo'n model de mogelijkheid tot *oordeelmatige kleuring van de voorspelling*. Eigenlijk heeft die inkleuring reeds plaats gevonden in het ontwerp van het causale economische model, met name in de specificatie van de gedragsvergelijkingen. In de voorspelfase wordt nog in andere zin oordeelmatige informatie in het model ingebracht, namelijk via autonome termen in de gedragsvergelijkingen die wel *informatiefactoren* (in het Engels 'constant adjustment' of 'add-factors') worden genoemd. Met behulp van zo'n informatiefactor geeft de voorspeller aan in hoeverre hij denkt dat het toekomstig verloop van een specifieke macro-economische grootte af zal wijken van het normale verloop. Op deze wijze kan informele en intuïtieve kennis in de modelvoorspelling worden ingebracht, maar ook formele kennis over gebeurtenissen waar het model geen rekening mee houdt kunnen op deze wijze op consistente wijze worden gecombineerd met de gedragsmatige beschrijving die het model van het verleden geeft. Bij zulke informatiefactoren valt bijvoorbeeld te denken aan de extra vraag naar televisietoestellen tijdens de Olympische Spelen of het intuïtieve gevoel dat het met de negatieve vermogens effecten van een beurscrisis (zoals die van oktober 1987) nogal mee zal vallen en dat deze crisis niet tot de conjuncturele inzinking zal leiden waar het model op basis van ervaringen uit het verleden op wijst.

Het gebruik van deze informatiefactoren voor het maken van macro-economische voorspellingen is reeds zeer oud. Al in 1953 werden ze toegepast bij het maken van voorspellingen met het 'Klein-Goldberger-model' van de Verenigde Staten. Kortom, de mogelijkheid om via deze informatiefactoren de voorspellingen te sturen, wordt wel als het grootste voordeel van het gebruik van causale macro-economische modellen gezien ten opzichte van de mechanische voorspelmethoden.

Het gebruik van informatiefactoren in relatie met macro-economische modellen dwingt de voorspeller om zijn inzichten over toekomstige ontwikkelingen die van het normale patroon afwijken, op een kwantitatieve manier te expliciteren. Een complicatie daarbij, die in de gedragsmatige beschrijving van de structuur van de economie door deze modellen besloten ligt, is dat informatiefactoren niet aan definitievergelijkingen mogen worden toegevoegd, doch uitsluitend aan gedragsvergelijkingen. Indien de voorspeller bijvoorbeeld gelooft dat zijn model een te lage voorspelling van het nationaal inkomen oplevert, is hij hierom verplicht aan te geven aan welke gedragsvergelijking hij een informatiefactor wil toevoegen om zijn voorspelling van het nationaal inkomen op te waarderen. Wanneer het model de gebruikelijke structuur in de traditie van Tinbergen heeft, wordt immers in het model het nationaal inkomen bepaald in een definitievergelijking als som van een aantal verschillende vraagfactoren. Bovendien dient de voorspeller zich ervan bewust te

zijn dat zo'n economisch structuurmodel uit een aantal simultane vergelijkingen bestaat en dat het toevoegen van een informatiefactor aan één gedragsvergelijking invloed heeft op de voorspelling van alle andere endogene variabelen van het model.

Dit is dan ook één van de belangrijkste voordelen van het gebruik van een macro-economisch model voor voorspellingen: de voorspellingen zijn *onderling consistent* en de veronderstellingen die aan de voorspellingen ten grondslag liggen worden hierdoor *expliciet* gemaakt. Zo dient de omvang van de informatiefactoren op zich plausibel te zijn. Een manier om deze plausibiliteit te beoordelen is de informatiefactoren te berekenen die in een dynamische modelsimulatie over het recente verleden de modeluitkomsten aan de realisaties gelijk doen zijn. Aldus geven deze endogene bepaalde autonome termen aan bij welke omvang van de informatiefactoren het model het recente verleden perfect zou hebben voorspeld.

4.1 Organisatie van voorspelrondes

Bij de instellingen die op professionele basis periodiek consistente voorspellingen voor de (wereld)economie opstellen zijn vanzelfsprekend vele verschillende deskundigen betrokken. Het voorspellen van zo'n groot aantal samenhangende economische grootheden wordt binnen iedere voorspelronde afgewikkeld in een iteratieve procedure. Een voorspelronde omvat daarbij een volledige cyclus van het opstellen van nieuwe voorspellingen, tot en met de publikatie ervan. Gesteld kan worden dat de kunst van het voorspellen met name op deze instellingen tot wasdom komt. Het groot aantal betrokkenen bij de voorspellingen brengt met zich mee dat de voorspelrondes goed georganiseerd dienen te zijn en volgens een strak en vooraf vastgesteld tijdschema verlopen.

In zo'n voorspelronde is het eerst van belang de toekomstwaarden van de exogene grootheden vast te stellen. Met name wordt daarbij aandacht besteed aan het voorgenomen beleid, dat voor de officiële instellingen zoals de OESO, maar ook het CPB als een gegeven geldt en dus als onweersproken uitgangspunt voor de veronderstelling, ook al hebben de deskundigen wel eens twijfel of de beleidsvoornemens daadwerkelijk zullen worden uitgevoerd. Op basis van deze veronderstellingen over de modelexogenen wordt vervolgens een eerste voorspelling met het model gemaakt. Aan de hand van deze uitkomst voegen vervolgens de deskundigen hun oordeelmatige informatie toe. Een nieuwe modeluitdraai, waarin deze oordeelmatige informatie is ingebracht, dient vervolgens om de verschillende en soms tegenstrijdige meningen van de deskundigen met elkaar in overeenstemming te brengen. Zo kan het voorkomen dat de landendeskundigen collectief te pessimistisch zijn over de betalingsbalansontwikkeling in 'hun' verschillende landen. In deze context betekent dit dat de wereldinvoer groter is dan de werelduitvoer, hetgeen natuurlijk niet mogelijk is. De betalingsbalansdeskundigen die verantwoordelijk zijn voor de consistentie op dit gebied, gaan dan met de landendeskundigen om de tafel zitten en vragen in welke mate deze geneigd zijn om de door hen voorspelde betalingsbalansposities op te waarderen. Aldus komt uiteindelijk de te publiceren voorspelling in een aantal stappen van hoor en wederhoor tot stand. De kunst van het voorspellen is daarbij vooral om te zorgen dat in dit communicatieproces de deskundigheid van alle betrokkenen zo goed mogelijk in de voorspellingen tot uitdrukking komt.

Het voorgaande maakt duidelijk dat er een zekere mate van spanning bestaat tussen de formele methodiek van het voorspellen en de kunst van het voorspellen waarin de menselijke intuïtie zo'n grote rol speelt. Samuelson (1975) illustreert deze spanning aan de hand van de volgende anekdote:

'Toen Robert Adams bij het Massachusetts Institute of Technology zijn proefschrift schreef over de nauwkeurigheid van verschillende voorspelmethoden bleek dat één van de beste methodieken indertijd een man was: Sumner Slichter. Vanuit wetenschappelijk oogpunt was het een correcte maar ook een droeve constatering. Slichter maakte namelijk zijn voorspellingen op basis van intuïtie en kon derhalve zijn voorspelkunsten niet aan een assistent of aan een jongere generatie van economen overdragen. Hij nam deze kundigheid mee in het graf, of verloor die misschien nog wel eerder, aan het eind van zijn leven.'

Het grote probleem dat Samuelson in het spanningsveld tussen wetenschappelijke voorspelmethoden en de kunst van het voorspellen aansnijdt is dat wetenschappelijke methoden reproduceerbaar zijn en daarmee volledig overdraagbaar zijn. De kunst van het voorspellen is daarentegen een persoonlijke bekwaamheid die slechts in zeer beperkte mate aangeleerd kan worden.

4.2 Kwaliteit van voorspellingen

Voor het gebruik van de voorspellingen in de beleidsanalyse is het natuurlijk van cruciaal belang om te weten hoe goed die voorspellingen zijn. Des te groter de reputatie van de voorspellingen, des te meer gewicht deze zullen hebben in de beleidsvoorbereiding. Anders dan de raadselachtige en vaak moeilijk interpreteerbare voorspellingen van waarzegsters laat een macro-economische voorspelling dat het nationaal inkomen van een land in het komend jaar met zeg 3% zal toenemen, weinig ruimte voor interpretatieverschillen. De kwaliteit van de voorspellingen kan daardoor achteraf worden beoordeeld aan de hand van de *voorspelfout*, d.w.z. het verschil tussen de voorspelling en de realisatie. Het gaat daarbij dus niet zozeer om het gebruik van de voorspelfout als criterium voor modelselectie maar als criterium voor modelkwaliteit.

Ook al is de voorspelling eenduidig, toch is de oordeelsvorming over de voorspelfout en daarmee over de modelkwaliteit niet zonder valkuilen. Dit geldt bijvoorbeeld wanneer men de kwaliteit van twee verschillende voorspelmethoden c.q. voorspellers met elkaar wil vergelijken. Een reden is dat deze voorspellingen betrekking kunnen hebben op verschillende definities van dezelfde grootheden. Of omdat de voorspellingen gemaakt zijn op verschillende tijdstippen. Over het algemeen zal immers de later gemaakte voorspelling beter zijn dan de op een vroeger tijdstip gemaakte voorspelling aangezien de voorspeller op het latere tijdstip over meer informatie kon beschikken. Nog een reden die het oordeel over de voorspelkwaliteit vaak vertroebelt, is dat er geen overeenstemming bestaat over de werkelijke waarde van de realisatie. Zo worden de definitieve cijfers van de Nationale Rekeningen pas na een aantal jaren gepubliceerd. Een bijkomstigheid bij de beoordeling van de kwaliteit van de voorspellingen van instanties zoals het CPB is dat deze in de beleidsanalyse een dermate zwaar gewicht in de schaal leggen, dat ze het karakter kunnen krijgen van hetzij zelfvernietigende, hetzij van zelfvervullende voorspellingen.

Ondanks deze moeilijkheden bestaan er toch veel studies waarin bij wijze van consumentenonderzoek de kwaliteit van verschillende voorspellingen met elkaar wordt vergeleken (zie Boeschoten en Van Els, 1994, voor een betrekkelijk recent voorbeeld voor Nederland). Hoewel volgens sommige vroegere studies mechanische voorspellingen op de korte termijn nauwelijks blijken onder te doen en soms zelfs beter blijken te scoren dan de met causale modellen gemaakte voorspellingen, laat het zich tegenwoordig toch aanzien dat de voorspellers die macro-economische structuurmodellen als hulpmiddel bij hun voorspellingen gebruiken, door de bank genomen betere voorspellingen af leveren dan de mechanische voorspellingen die zonder menselijke tussenkomst direct uit de computer rollen.

De belangrijkste kanttekening bij dit soort competitie is echter dat deze geen uitsluitsel bieden over het hoe en het waarom van de voorspelfouten. Het leert niet waarom sommige modellen of methoden beter zijn dan anderen. Anders gezegd, de uitslag van een voorspelcompetitie mag dan de winnaar een glorieus gevoel geven, deze geeft geen nuttige aanwijzing aan degene die de volgende wedstrijd wil winnen. En het laatste is nodig om de produktkwaliteit van de voorspellingen te verbeteren, hetgeen uit het oogpunt van de welvaart gewenst is.

4.3 Analyse van voorspelfouten

Daarom is het veel beter om de modellen en hun gebruik bij het maken van economische voorspellingen op een zodanige wijze te vergelijken dat we te weten komen wat de zwakke en de sterke punten van de voorspelprocedure zijn. In die zin is onderzoek gedaan naar de *informatie-inhoud* van mechanische voorspellingen met macro-economische modellen, en is deze vergeleken met de informatie-inhoud van voorspellingen met modellen waaraan oordeelmatige informatie is toegevoegd. Dergelijk onderzoek geeft niet alleen antwoord op de vraag in welke mate de voorspellers verschillende informatie gebruiken, of misschien zelfs alle beschikbare informatie, maar geeft ook een aanwijzing of het wellicht nuttig is de voorspellingen uit verschillende bronnen met elkaar te combineren. Zie hierover de eerdere opmerkingen over de mogelijkheid tot combinatie van voorspellingen.

In dit verband is ook onderzocht in hoeverre de oordeelmatige inbreng via de informatiefactoren de voorspelling die uit het model rolt verbetert. Volgens onderzoek van McNees (1990) blijkt dit inderdaad het geval te zijn, hoewel af en toe de mechanische voorspelmethode beter lijken te voldoen. Een interessante vraag daarbij is of de voorspellers teveel of te weinig op hun intuïtie afgaan en of derhalve de via de informatiefactoren gemaakte correcties op de mechanische voorspellingen over het algemeen te groot of te klein zijn. Het blijkt dat een aantal Amerikaanse voorspellers iets te sterk hun modellen corrigeert. De boodschap hierbij is dat de voorspellers inderdaad informatie dienen toe te voegen aan hun modelvoorspellingen maar dat zij dit met grote terughoudendheid moeten doen en tot op zekere hoogte best geloof mogen hechten aan hun modellen. Het blijft daarbij altijd van belang goed in het oog te houden of de omvang van de toegevoegde informatiefactoren plausibel is.

Om die reden is een zogeheten *post-mortem analyse* (in de terminologie van de OESO, zie Llewellyn, Porter en Samuelson, 1985) een essentieel onderdeel van de voorspelprocedure. Deze analyse poogt de bronnen c.q. de oorzaken van de voorspelfouten te achterhalen. De kennis over de oorzaken van de voorspelfouten is voor de voorspeller nuttig om zijn voorspellingen (en/of zijn voorspelmethode) te kunnen verbeteren. Voorspelfouten kunnen verschillende oorzaken hebben, namelijk:

1. specificatiefouten van het model;
2. schattingsfouten bij de coëfficiëntwaarden;
3. feitelijke gegevens die niet op het gehanteerde theoretische concept aansluiten;
4. het verkeerd inschatten van de ontwikkeling van de exogene variabelen, waaronder de beleidsvoornemens;
5. foutieve informatiefactoren en dus de verkeerde oordeelmatige aanpassingen van de modelvoorspellingen
6. het gebruik van voorlopige gegevens voor de vertraagde endogenen in het model.

In de praktijk blijkt het echter zeer moeilijk om in een post-mortem analyse al deze verschillende bronnen van voorspelfouten te onderscheiden. Zelfs in retrospect is het bijna onmogelijk om te bepalen of een voorspelfout veroorzaakt is door een verkeerde beschrijving in het model van het economisch gedrag, of door een preferentieverschuiving, waarmee de voorspeller via een informatiefactor rekening had moeten houden. Daar komt bij dat zelfs een volledig juiste voorspelling nog geen goede voorspelprocedure behoeft te impliceren aangezien deze voorspelling de resultante kan zijn van twee fouten die elkaar toevalligerwijze volledig opheffen.

Een post-mortem analyse is niet alleen nuttig om de macro-economische voorspeller te laten inzien waar hij gefaald heeft, maar ook omdat deze analyse inzicht verschaft in de economische ontwikkeling in het recente verleden. Indien men met het model een dynamische simulatie over het recente verleden uitvoert, met de gerealiseerde waarden van de exogenen en met realisaties voor de beginwaarden van de vertraagde endogenen, geven de *ex-post* voorspelfouten aan welke ontwikkeling in het recente verleden als bijzonder gekenschetst kan worden. Ondanks het grote belang uit analytisch oogpunt om de verschillende onderdelen van de voorspelfouten te ontrafelen, bestaat in de modelmatige beleidsanalyse vooralsnog geen vaste methodiek om de verschillende oorzaken van bijzondere en abnormale economische ontwikkelingen te classificeren. Dit is een gebied waar de wetenschap van het voorspellen, anders dan in de modelselectie en specificatie-analyse van tijdreeksmodellen, naar mijn mening nog wel een grote hulp aan de kunst van het voorspellen kan bieden.

5. Besluit

Deze bijdrage aan de afscheidsbundel van Merkies gaat over de methodologie en gedachtenvorming rond voorspellingen in de economie. Goede economische voorspellers dienen zowel te beschikken over econometrisch-technische vaardigheden als over kennis van economische mechanismen. Bovenal is echter een scherpe intuïtie nodig voor het belang van recente gebeurtenissen voor de toekomstige economische ontwikkeling. Econometristen zijn vooral geneigd om een mogelijke kwaliteitsverbetering van de voorspellingen te zoeken bij meer geavanceerde technieken. Het proefschrift van Merkies, waarin hij de keuze voor het model dat het beste kan voorspellen op mechanische wijze laat bepalen door de omvang van het voorspellingsinterval, valt binnen deze denkrichting en dateert uit een tijdperk toen technische verfijningen nog tot een verbetering van de voorspelkwaliteit konden leiden.

In deze bijdrage wordt de door Merkies gepropageerde methodiek van modelselectie in retrospect besproken en wordt deze gezien in het licht van andere denkbeelden over economische voorspellingen die intussen naar voren zijn gekomen. Daarbij is de aandacht verschoven van de econometrische naar de economische kant van de voorspelmedaille, of met andere woorden, van de techniek - of zo men wil van de wetenschap - van het voorspellen naar de kunst van het voorspellen. De indruk bestaat dat, zeker op het gebied van de macro-economische voorspellingen, met een verfijning van het technisch instrumentarium niet veel winst meer valt te behalen. Of deze winst wel kan worden verkregen via een betere theorievorming en een groter inzicht in de mechanismen die in de economie werkzaam zijn, moet zich nog bewijzen. Benutting van informatie uit micro-economische gegevensbestanden lijkt een optie om tot een kwaliteitsverbetering van de voorspellingen te komen, maar, zoals Gorter, Hassink en Pels (1996) aantonen, blijken bij voorbeeld bedrijven ook niet altijd hun toekomstige omzet en werkgelegenheid goed te kunnen voorspellen. In ieder geval lijkt het van belang om de analyse van voorspelfouten meer ter lering en minder ter vermaak te benutten. Hiermee bedoel ik dat het vergelijken van de voorspelprestaties van modellen en/of van de voorspellers die de modellen gebruiken aan de hand van de voorspelfouten, minder zinvol is dan een analyse van de achtergrond van die voorspelfouten. Wellicht kan het adagium dat men van zijn fouten dient te leren de kunst van het voorspellen in beweging houden.

Literatuur

Boeschoten, W.C. en P.J.A. van Els (1994), 'De voorspelkwaliteit van macro-economische modellen: een vergelijkende analyse voor het DNB-model MORKMON', *Maandschrift Economie* 58, 377-387.

Box, G.E.P. en G.M. Jenkins (1970), *Time Series Analysis*, Holden-Day, San Francisco.

Butter, F.A.G. den (1975), 'Prediction of two periods by simple autoregressive models with one lag', *De Economist* 123, 58-74.

Butter, F.A.G. den (1976), 'The use of monthly and quarterly data in an ARMA model', *Journal of Econometrics* 4, 311- 324.

Butter, F.A.G. den (1992), 'Scope and limitations of statistical methods for macroeconomic forecasting', *Kwantitatieve Methoden* 13, nr. 40, 5-21.

Christoffersen, P.F. en F.X. Diebold (1996), 'Further results on forecasting and model selection under asymmetric loss', *Journal of Applied Econometrics* 11, 455-473.

Compaijen, B. en F.A.G. den Butter (1996), *De Nederlandse Economie, Beschrijving en Besturing* (5e druk), Wolters-Noordhoff, Groningen.

Gameren, E. van, D.E.G. Molenaar en R. Paap (1993), 'Evaluatie van modelselectie met voorspelcriteria', *Kwantitatieve Methoden* 14, nr. 44, 45-57.

Gorter, C., W. Hassink en E. Pels (1996), 'Werkgelegenheidsvoorspellingen van bedrijven', *Tijdschrift voor Arbeidsvraagstukken*, 1996-12, nr. 3, 264-272.

Granger, C.W.J. (1996), 'Can we improve the perceived quality of economic forecasts?', *Journal of Applied Econometrics* 11, 455-473.

Keuzenkamp, H.A. (1995), 'The econometrics of the Holy Grail': a critique, *CentER Discussion Paper* nr. 9521.

Kloek, T. (1974), 'Boekbespreking A.H.Q.M. Merkies: Modelkeuze via voorspellingsintervallen', *De Economist* 122, 68-72.

Kool, J.T.C. (1989), *Multi-step Prediction Models for Stationary Multivariate Time Series*, Free University Press, Amsterdam

Llewellyn, G.E.J., S.J. Porter en L.W. Samuelson (1985), *Economic Forecasting and Policy: the International Dimension*, Routledge and Kegan Paul, Londen.

Manning, A. (1987), 'The integration of trade union models in a sequential bargaining framework', *Economic Journal* 97, 121-139.

McNees, S.K. (1990), 'Man vs. model? The role of judgement in forecasting', *New England Economic Review*, July/August 1990, 41-52.

Merkies, A.H.Q.M. (1972), *Modelkeuze via Voorspellingsintervallen*, Academic Service, Amsterdam.

Merkies, A.H.Q.M. (1973), *Selection of Models by Forecasting Intervals*, Reidel Publishing Company, Dordrecht/Boston.

Samuelson, P.A. (1975), 'The art and science of macromodels over 50 years', in G. Fromm and L.R. Klein (eds.), *The Brookings Model, Perspective and Recent Developments*, North-Holland, Amsterdam, 3-10.

Sims, C.A. (1980), 'Macroeconomics and reality', *Econometrica* 48, 1-48.

Wallis, K.F. (1989), 'Macroeconomic forecasting: a survey', *Economic Journal* 99, 28-61.

Samenvatting

In deze bijdrage wordt het idee uit het proefschrift van Merkies om de modelkeuze te laten afhangen van de omvang van het voorspellingsinterval in retrospect tegen de achtergrond van meer recente denkbeelden op het gebied van economische voorspellingen besproken. Daarbij valt een verschuiving van de modelkeuze op basis van econometrisch-technische criteria naar modelgebruik op grond van economische plausibiliteit en intuïtie waar te nemen.